EXHIBIT 9

J.W. IZARD & D.A. KENDALL

MOLECULAR MICROBIOLOGY 13 765-773, 1994

EXHIBIT 9

Molecular Microbiology (1994) 13(5), 765–773

MicroReview

Signal peptides: exquisitely designed transport promoters

Jennifer W. Izard and Debra A. Kendall*
Department of Molecular and Cell Biology, The University of Connecticut, Storrs, Connecticut 06269, USA.

Summary

Prokaryotic proteins destined for transport out of the cytoplasm typically contain an *N*-terminal extension sequence, called the signal peptide, which is required for export. It is evident that many secretory proteins utilize a common export system, yet the signal sequences themselves display very little primary sequence homology. In attempting to understand how different signal peptides are able to promote protein secretion through the same pathway, the physical features of natural signal sequences have been extensively examined for similarities that might play a part in function. Experimental data have confirmed statistical analyses which highlighted dominant features of natural signal sequences in *Escherichia coli*: a net positive charge in the *N*-terminus increases efficiency of transport; the core region must maintain a threshold level of hydrophobicity within a range of length limitations; the central portion adopts an α-helical conformation in hydrophobic environments; and the signal cleavage region is ideally six residues long, with small side-chain amino acids in the −1 and −3 positions. This review focuses on the parallels between signal peptide physical features and their functions, which emerge when the results of a variety of experimental approaches are combined. The requirement for each property may be ascribed to a potential interaction that is critical for efficient protein export. The summation of the key physical features produces signal peptides with the flexibility to function in multiple roles in order to expedite secretion. In this way, nature has indeed evolved exquisitely tuned signal sequences.

Received 29 March, 1994; accepted 30 May, 1994. * For correspondence: E-mail kendall@uconnvm.uconn.edu; Fax (203) 4861784.

Introduction

Proteins destined for extracytoplasmic compartments in both eukaryotic and prokaryotic cells typically contain an *N*-terminal extension sequence called the signal peptide or leader sequence, which plays a crucial role in increasing the efficiency of protein transport across membranes. Although signal peptides lack primary sequence homology (Watson, 1984), they display similar physical characteristics. It is well established that the shared features enable different signal peptides to interact with common elements in a 'general secretory pathway' in *Escherichia coli* (for a review see Pugsley, 1993). Furthermore, since signal peptides can be interchanged between eukaryotic and prokaryotic systems, parallels between the two transport pathways are likely to exist.

Examination of natural signal sequences has produced a description of the typical signal peptide in *E. coli*: the *N*-terminus is five to six residues long and has, on average, two positively charged amino acids; the hydrophobic core is about 12 residues long, contains highly hydrophobic residues and is uncharged; the *C*-terminus is six residues long, has neutral small side-chain amino acids at positions −1 and −3, and often includes a proline or glycine residue (Perlman and Halvorson, 1983; von Heijne and Abrahmsén, 1989). Numerous investigations have confirmed the importance of several of the features revealed by statistical evaluations and have begun to disclose information as to why these properties are required. Protein secretion entails the successful completion of a series of steps: synthesis, membrane insertion, translocation, signal peptide recognition and cleavage by leader peptidase, and final localization of the mature protein. The signal peptide may play multiple roles in accomplishing these steps by retarding the folding of the precursor (Weiss and Bassford, 1990; Hardy and Randall, 1991) and through specific interactions with the lipid and protein components of the transport pathway.

Genetic selection experiments and biochemical studies have been employed to identify proteinaceous components of the bacterial export pathway (for reviews see Schatz and Beckwith, 1990; Mizushima *et al.*, 1991; Wickner *et al.*, 1991). SecB is a cytoplasmic protein which acts as a molecular chaperone and helps maintain a transport-competent state for some exported proteins (Hardy and

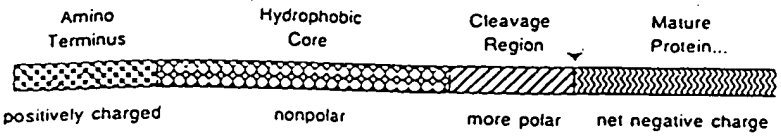

Fig. 1. Characteristics of a prokaryotic signal peptide. Cleavage of the signal peptide occurs at the position marked by the arrow. In addition to the features of the signal peptide, the diagram includes the early mature portion of the transported protein, which typically carries a neutral or net negative charge.

Randall, 1991). SecA is an ATPase found in two states, as a soluble and as a membrane-associated protein bound on the cytoplasmic side of the inner membrane (for a review see Oliver, 1993). SecD, SecE, SecF and SecY are integral cytoplasmic membrane proteins. Recently, SecA and SecY have been shown to crosslink to a translocating polypeptide chain (Joly and Wickner, 1993). SecE and SecY directly interact and seem to be involved in membrane translocation (for a review see Ito, 1992; Taura *et al.*, 1993). SecD appears to be important in the release of the mature protein from the inner membrane (Matsuyama *et al.*, 1993), while the leader peptidases I and II are responsible for the actual signal peptide cleavage step (for a review see Dalbey, 1991). In addition, Fth has been implicated as part of a ribonucleoprotein signal-recognition particle in *E. coli* (Luirink *et al.*, 1992; Phillips and Silhavy, 1992) and p12 stimulates protein translocation in a reconstituted system (Nishiyama *et al.*, 1993). Progress in discovering the integral components of the export pathway has led to the proposal of models for the assembly and function of the translocation complex. We can now begin to combine our knowledge of signal peptides and the components of the export pathway in order to elucidate the key interactions involved in directing a protein to its final destination.

The distribution of charged residues

Secretory proteins are partially characterized by the pattern of charged residues in the signal sequence and the *N*-terminus of the mature protein. Typically, the signal peptide has a net positive charge in the *N*-terminus and lacks charged residues in the core region and *C*-terminus, although there can be a slightly higher incidence of negatively charged amino acids near the cleavage site. The positive charge requirement has been shown to be independent of residue type; analysis of signal peptide mutants indicates that lysine and arginine residues can be interchanged, and even a histidine at the proper pH provides the positive charge required for translocation (Sasaki *et al.*, 1990). Efficient transport is, however, dependent on the location of these residues and the introduction of positive charges elsewhere in the signal peptide or in the *N*-terminus of the mature protein is not tolerated. There are several possible explanations for this bias: a positive charge beyond the *N*-terminus may alter the charge dipole across the signal peptide, provide competition with the *N*-terminal positive charge for an important electrostatic interaction (MacIntyre *et al.*, 1990), or simply disrupt the features desired in another region (e.g., the hydrophobicity in the core).

Von Heijne (1986) recognized the 'dipolar' character of bacterial signal sequences as a potentially important feature for function (Fig. 1). The distribution of positively and negatively charged residues at opposite ends of the signal peptide as well as the regular alignment of carbonyl and amino groups along the backbone of an α-helical region contribute to the net positive dipole. Although the extent to which the dipole moment contributes to transport efficiency is not well understood, it is plausible that the membrane potential of the prokaryotic inner membrane may facilitate insertion of a signal peptide with this overall charge distribution (for reviews see Geller, 1991; Driessen, 1992); the energetics of the protonmotive force might assist in the transport of negatively charged residues, while it would resist the transport of positively charged residues. In support of this theory, the 'positive-inside rule', predicting that hydrophobic sequences will be aligned in the membrane such that the most positively charged terminus is oriented toward the cytoplasm, accurately describes both transmembrane segments and cleavable signal peptides (Boyd and Beckwith, 1990; von Heijne, 1992).

Mutations that produce a net negative or zero charge at the *N*-terminus of the signal peptide result in considerably decreased rates of export (Inouye *et al.*, 1982; Vlasuk *et al.*, 1983; Puziss *et al.*, 1989). The negatively charged mutants temporarily accumulate in the cytoplasm but do eventually gain entry into the secretory pathway in a post-translational manner. Therefore, for wild-type precursors, the positive charge increases the efficiency of transport although it may not be absolutely required for secretion.

The significance of the positive charge at the *N*-terminus is often attributed to the need for an initial electrostatic attraction to the negatively charged phospholipids of the inner membrane, which can be followed by insertion of the hydrophobic region of the signal peptide into the bilayer or a proteinaceous component therein, e.g. the loop model (Inouye and Halegoua, 1980; for a review see de Vrije *et al.*, 1990). Upon signal peptide interaction with the membrane, the *N*-terminal positive charge may also facilitate initial loop formation by interacting with the negatively charged residues that often exist around the signal peptide cleavage site (Bosch *et al.*, 1989).

Formation of such a loop was implicated by the biophysical analysis of the conformation and orientation of a synthetic PhoE signal peptide in a lipid environment (Batenburg *et al.*, 1988). Indeed, anionic phospholipids seem to promote an association of synthetic PhoE signal sequences with membrane vesicles (Keller *et al.*, 1992) and a reduced amount of anionic lipid results in a decrease in the translocation ability of prePhoE *in vivo* (deVrije *et al.*, 1988) and of SecA-independent M13 procoat *in vitro* (Kusters *et al.*, 1994). Efficient translocation of a hybrid of outer membrane protein F and the major lipoprotein of *E. coli* (OmpF–Lpp) *in vitro* is dependent on both the charge of the signal sequence and the phosphatidylglycerol content of the membrane vesicles (Phoenix *et al.*, 1993b). Doxorubicin, a positively charged antibiotic, and polylysine inhibit translocation of OmpF–Lpp into vesicles by preferentially binding phosphatidylglycerol and thereby preventing the required interaction (Phoenix *et al.*, 1993a). Interestingly, lengthening the core region, and thus increasing hydrophobicity, allows translocation in the presence of these competitors, regardless of the *N*-terminal charge.

In addition to a favourable interaction with the phospholipids, the positive charge may foster an interaction with a component of the export pathway. Crosslinking of purified SecA to the mature portion of a precursor protein was found to be dependent on the net positive charge at the *N*-terminus of the precursor (Akita *et al.*, 1990). Furthermore, mutations in the *prlD* (SecA) gene partially suppress a secretion deficiency caused by lowering the charge at the *N*-terminus (Puziss *et al.*, 1989). Although suppression may imply physical contact between the signal peptide and the secretory component, no direct evidence for such an interaction is available. The improved secretion may instead be the indirect consequence of a conformational change in SecA which loosens the requirements for entrance into the secretory machinery. This kind of effect could explain a similar observation of limited secretion of alkaline phosphatase (PhoA) lacking a signal peptide in a *prlA* (SecY) mutant strain (Derman *et al.*, 1993).

Hydrophobicity profile

The occurrence of several hydrophobic amino acids, clustered in the core region, is often regarded as the hallmark of a signal peptide. The importance of this signature is underscored by the observation that a variety of mutations in this central domain produce defective signal peptides (for a review see Gennity *et al.*, 1990). For example, disruption of the core region with a charged residue usually prevents normal secretion (Bankaitis *et al.*, 1984; Stader *et al.*, 1986; Kendall *et al.*, 1990). Introduction of a neutral but polar group, such as asparagine, has a similar effect (Goldstein *et al.*, 1991), and a range of different deletion mutations within this region results in export defects (Emr and Silhavy, 1983; Lehnhardt *et al.*, 1987).

While the effect of mutations in the hydrophobic core region is readily scored, the reason a given mutation produces its effect is less easily evaluated. The effect of single amino acid substitutions can be dependent on the position and the nature of the change; this is particularly apparent in mutations at the periphery of the core where a hydrophilic residue may be acceptable as long as a hydrophobic stretch remains intact (von Heijne, 1986). The analysis is complicated by the variety of amino acids in these signal sequences, the contribution of varying types of conformation and the different degrees of side-chain hydrophobicity. Furthermore, it is easy to overlook the bias of the experimental approach used to generate a given hydropathy table which may be used to analyse the mutations (von Heijne, 1985a).

In order to independently address questions regarding length, hydrophobicity and conformation experimentally, it is often advantageous to eliminate some of the variables common to natural signal sequences by replacing segments with polymers of only one or a few different amino acid residues. The first application of this method involved multiple amino acid changes to produce a PhoA signal peptide with a polyleucine-containing core region (Kendall *et al.*, 1986). This mutant was processed at least as efficiently as the wild type, emphasizing the importance of physical properties over primary sequence and establishing that rationally designed, multiple residue changes can be introduced into a signal peptide with retention of biological activity. The approach is simplified through the use of cassette mutagenesis (Kendall and Kaiser, 1988) and has been employed to systematically study the hydrophobic character of the signal peptide.

By replacing the natural core region of the PhoA signal peptide with homopolymers of leucine, valine, or alanine; the interrelationship between the hydrophobicity and length of the signal sequence was examined (Chou and Kendall, 1990). While retaining the natural length of the core segment, the extent of processing and transport declined in these mutants in parallel with a decrease in the hydrophobicity of the constituent residue used. The core region that is composed of 10 leucine residues supported very rapid and complete transport while that containing 10 valine residues was only weakly exported and the 10 alanine core was severely defective. However, the loss in transport activity can be partially compensated for by increasing the length of the hydrophobic polymer used. Thus, a high net hydrophobicity contributed by increased length can partly restore function diminished by a lower mean hydrophobicity per residue. However, extremes in length are not readily tolerated: it was found that a core region of 20 leucine residues produced a translocated yet anchored sequence, suggesting that length

can be critical for topological alignment for cleavage. Lower limits on the core segment length have also been documented for which even the most hydrophobic sequences do not provide a functional signal peptide (von Heijne, 1985b; Chou and Kendall, 1990; Hikita and Mizushima, 1992a).

As increased length does not completely compensate for the core segments with less hydrophobic residues, the ideal sequence will meet a threshold 'hydrophobic density' to ensure rapid translocation, i.e. a high mean hydrophobicity per residue must be attained within the absolute bounds of length limitations. An estimation of that threshold level was accomplished by 'titrating' a non-functional polyalanine-containing signal peptide with the more hydrophobic residue leucine (Doud *et al.*, 1993). Using precursor processing as an earmark of transport, a clear non-linear dependence on signal peptide hydrophobicity was observed. The sigmoidal curve generated from the data is consistent with a simple model representing a two-state equilibrium between untransported and transported species. The equilibrium, for any one mutant, is defined by the number of alanine and leucine residues in its signal peptide core region. For those mutants with few leucine residues, the free energy change associated with transport is positive and the equilibrium is far to the left. With greater numbers of leucine residues, the free energy change becomes negative and the equilibrium becomes favourable for transport. The midpoint of the curve falls between alanine to leucine ratios of 6:4 and 5:5. The mutant cores with hydrophobicity just below this threshold exhibit significant precursor processing over time, suggesting that the free energy of these precursors is close to that required to drive the transition from an untransported to a transported state. Interestingly, the hydrophobicity of natural core regions is similar to this experimentally derived threshold level. Of 11 wild-type sequences for periplasmic proteins, only two contain core segments with a mean hydrophobicity significantly below the threshold, and no core regions are more hydrophobic than a mutant core composed of the equivalent of three alanines and seven leucines. In nature, there appears to be no advantage to evolving more highly hydrophobic signal sequences. Nevertheless, a handful of such sequences have been generated experimentally, including polymers of leucine (Kendall *et al.*, 1986), isoleucine (Kendall and Kaiser, 1988), phenylalanine (Rusch and Kendall, 1992) and sequences of mixed amino acid composition (Goldstein *et al.*, 1990); these have all proved more efficient than their wild-type counterparts for at least one step during transport.

While the hydrophobic nature of signal peptides has long been recognized as critical to the transport process, the role(s) that hydrophobicity plays has proved considerably more difficult to elucidate. Energetic considerations led to the formulation of 'the helical hairpin hypothesis' (Engelman and Steitz, 1981) and other thermodynamic models (von Heijne and Blomberg, 1979; Wickner, 1979), which suggest that the hydrophobic signal peptide fosters the spontaneous insertion of the secreted polypeptide into the membrane and that this interaction promotes the initiation of translocation across the bilayer. These theories rely on the free energy potential of the hydrophobic core to enable unassisted partitioning into the hydrophobic interior of the bilayer. Many studies using synthetic peptides and model membrane systems have confirmed that the signal peptide can readily insert into the acyl chain region of bilayers (Batenburg *et al.*, 1988; McKnight *et al.*, 1991). The ability of synthetic peptides to insert into membranes and the *in vivo* function of the corresponding signal sequences has further been correlated with the mean hydrophobicity of their core regions (Hoyt and Gierasch, 1991).

In addition to its role in membrane association, the hydrophobicity of the signal peptide has also been implicated in the translocation process; precursors containing hydrophobic core mutations have been identified that are membrane associated but not translocated (Thom and Randall, 1988; Chou and Kendall, 1990; Rusch and Kendall, 1992). The hydrophobicity of these mutants may fall sufficiently short that the lipid perturbations required for passage of the polypeptide chain do not occur, or perhaps a critical protein–hydrophobic core interaction involved in translocation cannot be achieved.

There is no direct evidence that the signal peptide hydrophobic core interacts with any of the protein components of the transport pathway. However, the hydrophobic unit may well serve as an important recognition factor for a protein-specific interaction. While the involvement of the primary sequence must be ruled out, specificity can be contributed by a composite of physical properties: upper and lower limits on length, propensity for α-helix formation (see below), exclusion of charged residues and requirements for a threshold level of hydrophobicity. Such an interaction, involving structural elements, has been suggested for recognition of polypeptide chains by heat-shock proteins (Landry and Gierasch, 1991; Gething and Sambrook, 1992) as well as for binding of hydrophobic portions of amphiphilic helices by calmodulin (O'Neil and DeGrado, 1990). Mutations in the *prlA* (SecY) gene (Puziss *et al.*, 1992; Francetic *et al.*, 1993; Olsen *et al.*, 1993), the *prlD* (SecA) gene (Fikes and Bassford, 1989), and the *prlG* (SecE) gene (Stader *et al.*, 1989) have been found which suppress mutations in the hydrophobic core of the signal peptide. Moreover, characterization of a collection of suppressor mutations in SecY reveals that these mutations are often clustered in distinct topological domains of SecY (Osborne and Silhavy, 1993). The clustering of key residues is compatible with the notion that these domains

constitu[illegible]inding sites, at least one of which may be involved in signal sequence interactions. SecA–hydrophobic core interactions are implicated by other experiments in which cells are treated with sodium azide, a specific and potent inhibitor of SecA function (Oliver *et al.*, 1990). Analysis of mutant precursors that are processed rapidly reveals that their transport becomes less and less affected by sodium azide as the signal peptide hydrophobicity is increased (Rusch *et al.*, 1994). The apparent decrease in sensitivity to sodium azide may actually reflect an increased affinity for SecA, and these higher affinity signal peptides may more efficiently utilize the few available target sites which remain unaffected by the inhibitor. In parallel with this trend, the mutants with highly hydrophobic signal peptides interfere with the transport of wild-type β-lactamase, suggesting that these mutants may outcompete other precursors for a component of the transport pathway.

Preferences for conformation

The secondary-structure requirements of a functional signal peptide are probably the most difficult to assess; this is because of the difficulty of measuring experimentally the structural content in a physiologically relevant state and the lack of reliable secondary-structure prediction schemes. Chou–Fasman (1978) analysis of natural signal sequences predicts that some have a propensity for forming α-helices while others have β-sheet potential (Austen, 1979). However, this analysis is based on the frequency with which amino acid residues exist in a particular structure in globular proteins, and it may not be appropriate for application to polypeptides which are potentially transfixed in the membrane (Wallace *et al.*, 1986). Indeed, recent spectroscopic studies of model peptides demonstrate that the α-helicity of a peptide in a membrane is largely governed by the segmental hydrophobicity of the peptide rather than the intrinsic helical propensities of the component residues (Li and Deber, 1993). This supports the proposal for α-helical formation during the transport of a functional signal peptide containing a decaisoleucine core region (Kendall and Kaiser, 1988).

Biophysical studies of synthetic signal peptides have revealed the existence of both β-sheet and α-helices depending on the environment (for a review see Jones *et al.*, 1990). In aqueous solution, signal peptides usually exhibit large amounts of random coil (Briggs and Gierasch, 1984; McKnight *et al.*, 1989) although aggregates of β-sheets have been observed (Batenburg *et al.*, 1988; Hoyt and Gierasch, 1991). Beta-structure is also apparent when synthetic signal peptides associate with the headgroups of phospholipid monolayers (Briggs *et al.*, 1986; Cornell *et al.*, 1989). However, in non-polar environments, such as trifluoroethanol, phospholipid vesicles and sodium dodecylsulphate micelles, α-helix structure is typical (Batenburg *et al.*, 1988; McKnight *et al.*, 1989). Remarkably, the α-helical content is often of the order of 70%, with the hydrophobic core offering the most stable helix but with some propagation of the α-helix through the *N*- and *C*-terminal segments (Bruch *et al.*, 1989; Bruch and Gierasch, 1990; Rizo *et al.*, 1993). Although the formation of an α-helical unit in and of itself is not sufficient for function (McKnight *et al.*, 1989; Hoyt and Gierasch, 1991), the stability of the helix in LamB and OmpA mutants has been correlated with *in vivo* function (Bruch and Gierasch, 1990; Rizo *et al.*, 1993). One advantage to helix formation is that the overall hydrophobicity is enhanced by the effective shielding of the polar amide groups via intrastrand hydrogen bonding, and such a hydrophobic α-helix may provide an important recognition element (Kendall *et al.*, 1986).

Sequence analysis of prokaryotic signal peptides reveals that they often contain a proline or another turn-promoting residue (e.g. glycine or serine) at the core-cleavage region boundary (von Heijne, 1986), but the role of this residue remains an unresolved issue. The presence of one of these residues has been proposed to introduce a β-turn, break the core region helicity, or lower the hydrophobicity, ultimately promoting recognition and cleavage by the signal peptidase. Mutagenesis studies have suggested that the efficiency of precursor processing declines as the cleavage region decreases in β-turn probability by Chou–Fasman parameters (Vlasuk *et al.*, 1984; Kadonaga *et al.*, 1985; Kuhn and Wickner, 1985; Duffaud and Inouye, 1988; Shen *et al.*, 1991). The analysis is complicated, however, because some of the mutants involve deletion of residues, thus shifting the location of the critical −1 and −3 residues (see below); some include charged residues, others involve residues in the mature portion of the protein, and yet others interchange two turn-promoting residues with widely different results. In contrast, NMR studies reveal that a synthetic LamB signal sequence can propagate some helical conformation in this region even in the presence of a glycine residue (Bruch *et al.*, 1989). Analyses of model peptides indicate that glycine (a known helix-breaker in globular proteins; Chakrabartty *et al.*, 1991) can be accommodated into a hydrophobic α-helix as readily as alanine (a strong helix-promoter; Lyu *et al.*, 1991) in non-polar environments (Li and Deber, 1993). These studies argue that the rules which govern soluble protein structure may not apply to membrane-interactive polypeptides, which can form helices even though they often contain a preponderance of residues traditionally considered to be helix destabilizing (Deber *et al.*, 1986). An idealized signal peptide, which lacks proline and contains only the helix-fostering residues leucine and alanine, efficiently exports PhoA (Laforet and Kendall, 1991). This result suggests that a clear demarcation in

polarity between the hydrophobic core and the cleavage region may be involved and that neither a proline nor a glycine is absolutely required. A recent statistical evaluation of natural signal sequences supports this notion (Schneider *et al.*, 1993). Additional methods of analysis seem critical in order to provide a balanced assessment of the conformation of this region.

It is important to consider that the formation of defined conformations could impose limitations on the length of the signal peptide which may, in turn, place topological constraints on its alignment in a membrane environment. For example, if the signal peptide consists of an α-helical core region of 10 residues and an extended *C*-terminus of six residues, the combined length of these regions would be about 3.6 nm. Assuming the *N*-terminal positive charge is anchored on the inner membrane surface, such a peptide arrangement could not span a typical membrane thickness (6 nm); it could span the distance of the acyl chains of the bilayer (around 3 nm) and place the cleavage site in the hydrophilic headgroups. The introduction of a β-turn in the signal peptide would further shorten the signal peptide. Only an entirely extended signal sequence could provide the length required to present the cleavage site in the periplasm, unless other accommodations are made. For example, the positive charge may not remain bound to the cytoplasmic phospholipid headgroups and instead may be drawn into the interior of the membrane. It is also possible that the bilayer thins at the location of the signal peptide (Tahara *et al.*, 1992) or that changes in conformation may bring the active site of leader peptidase into the interior of the bilayer during the transport process. Regardless of this, the conformations of the signal peptide must be compatible with its topological orientation with the membrane and the protein machinery therein.

Primary sequence specificity

Signal peptides generally contain very few primary sequence constraints. The exceptions are the amino acids at positions −3 and −1 upstream of the cleavage site. These residues constitute the recognition site of the leader peptidase enzyme and usually have small neutral side chains, such as alanine, glycine, serine and threonine, with alanine found especially frequently, i.e. the '−3, −1 rule' (von Heijne, 1984). Mutational analysis of the maltose-binding protein (MBP) cleavage region confirms these requirements although it indicates that cysteine is also allowed at the −3 and −1 positions while threonine at −1 strongly reduces processing (Fikes *et al.*, 1990). In several studies, the replacement of the −1 residue with an unfavourable amino acid can induce processing at an alternate site (Fikes *et al.*, 1990; Nothwehr and Gordon, 1990; Laforet and Kendall, 1991). It has also been demonstrated that proline at position +1 produces a non-cleavable substrate that competitively inhibits the leader peptidase activity (Barkocy-Gallagher and Bassford, 1992; Nilsson and von Heijne, 1[illegible]). In addition to the requirement for specific amino acids at −3 and −1, the location of these residues relative to the *C*-terminal end of the hydrophobic core is important, i.e. four and six residues, respectively, from the core-cleavage region boundary (Perlman and Halvorson, 1983; von Heijne, 1984; Schneider *et al.*, 1993; Jain *et al.*, 1994).

Combined features

For optimal signal sequence function, there is a requirement for a combination of physical features to create a signal sequence with the desired activity for each individual secretory protein. Overlapping qualities in a potentially optimal signal sequence have already been discussed: namely, the ability of a helical structure to enhance the overall hydrophobicity as well as the net dipole moment of the segment involved. Experimentally, the negative effects of altering one feature can often be attenuated by improving the quality of another property. For instance, the addition of a positive charge in the core region of lipoprotein signal peptide has been shown to partially suppress the effects of mutations created by a negatively charged *N*-terminus (Sung *et al.*, 1992); this suggests that the placement of the positive charge may not be critical for efficient transport as long as the core remains sufficiently hydrophobic for translocation. Another study demonstrated that enhanced hydrophobicity of the core regions of MBP signal peptide mutants can compensate for a decreased positive charge in the *N*-terminus (Puziss *et al.*, 1989). The offset of a suboptimal *N*-terminus is also manifest in a model secretory protein containing a core region of ample length and hydrophobicity; a core composed of nine leucine residues supports *in vitro* translocation of proOmpF–Lpp even with zero charge in the *N*-terminus. In contrast, a hydrophobic stretch of only seven or eight leucines requires a positively charged *N*-terminus for efficient translocation (Hikita and Mizushima, 1992b).

Optimized core regions can also overcome defects in the early portion of the mature protein. Replacement of six residues in the *N*-terminal region of the mature protein with a homopolymer of serines creates a transport-incompetent PhoA precursor. This mutant shows little transport improvement when several wild-type residues are reintroduced, yet function is restored by incorporating a hydrophobic core of 10 leucine residues in the signal peptide (Rusch and Kendall, 1994). The results demonstrate that the high hydrophobicity of the signal peptide can override potential requirements for β-turn promoting and negatively charged residues that often exist in the mature portion of the protein. Lengthening the hydrophobic core of a proOmpA mutant can also compensate for the inhibitory

effect of positive charges in the early mature protein (MacIntyre *et al.*, 1990). Furthermore, MacIntyre examined natural precursor proteins and proposed that the unusual presence of a positive charge in the *N*-terminus of the mature protein is often preceded by a signal peptide which lacks disruptive proline or glycine residues. This implies that the frequent occurrence of these β-turn-promoting residues in natural signal sequences is not required but rather is tolerated if other features of the signal peptide are optimized for transport.

Optimized sequences: helpful or harmful?

Many examples in the preceding section show that an optimized core region can compensate for a defect in or around the signal peptide, suggesting that the functions of the different segments are not entirely distinct. The experimental evidence confirms the balance of features that exist in natural signal sequences, but it also poses several questions. Why are wild-type core regions not even more hydrophobic? Why do they not adopt more regular α-helices and evolve a more specific length? Studies with idealized sequences, highly hydrophobic homopolymers designed to minimize structural strains caused by varying amino acid composition, may provide a clue. Signal peptide mutants containing nine or 10 leucine residues, by some criteria, function better than the wild-type core. However, these core region mutants cause precursor accumulation of another wild-type protein in the cell, β-lactamase (Rusch *et al.*, 1994). This observation discloses an uneven competition that is created by the presence of a precursor with a superior signal peptide. Thus, although a highly hydrophobic core region apparently enhances its own export, it may disturb the balanced secretion of other exported proteins. This phenomenon would not occur if all signal peptides were excessively hydrophobic, but there is no apparent driving force for evolving signal sequences that are any more efficient. Furthermore, few combinations of amino acid residues could produce a hydrophobic segment equivalent to 10 leucines. Isoleucine and phenylalanine would more than suffice, but both may present structural distortions if used frequently, and the repeated use of a small subset of amino acids might tax the components used for protein synthesis. It is evident from the large array of amino acid combinations used in natural signal sequences (Watson, 1984) that the possible combinations which produce segments somewhat less hydrophobic than 10 leucines are not limiting.

Experiments involving idealized mutants suggest a high affinity of the highly hydrophobic and α-helical signal peptides for SecA. The possibility of binding one component too tightly may be counterproductive in a system which may rely on a transient recognition by each component in a series of steps ultimately leading to secretion. While we may discover means for increasing export by optimizing the hydrophobic core region or other segments, changes in this direction may not be beneficial to the overall health of the cell. Natural signal peptides are designed to incorporate a combination of features, creating a well-balanced unit. The amino acid variation within signal sequences allows flexibility for the multiple interactions that may be necessary to promote secretion of the attached protein without encumbering the pathway. In this light, natural signal peptides may be perceived as containing intentional flaws and wobbles (in otherwise uniform structures) which deter overuse of the system; nature has, indeed, evolved signal sequences which are precisely tuned for the transport process.

Acknowledgements

Research in the authors' laboratory is supported by The National Institutes of Health and The Connecticut Department of Economic Development.

References

Akita, M., Sasaki, S., Matsuyama, S., and Mizushima, S. (1990) *J Biol Chem* 265: 8164–8169.
Austen, B.M. (1979) *FEBS Lett* 103: 308–313.
Bankaitis, V.A., Rasmussen, B.A., and Bassford, Jr, P.J. (1984) *Cell* 37: 243–252.
Barkocy-Gallagher, G.A., and Bassford, Jr, P.J. (1992) *J Biol Chem* 267: 1231–1238.
Batenburg, A.M. Brasseur, R., Ruysschaert, J.-M., van Scharrenburg, G.J.M., Slotboom, A.J., Demel, R.A., and de Kruijff, B. (1988) *J Biol Chem* 263: 4202–4207.
Bosch, D., de Boer, P., Bitter, W., and Tommassen, J. (1989) *Biochim Biophys Acta* 979: 69–76.
Boyd, D., and Beckwith, J. (1990) *Cell* 62: 1031–1033.
Briggs, M.S., and Gierasch, L.M. (1984) *Biochemistry* 23: 3111–3114.
Briggs, M.S., Cornell, D.G., Dluhy, R.A., and Gierasch, L.M. (1986) *Science* 233: 206–208.
Bruch, M.D., and Gierasch, L.M. (1990) *J Biol Chem* 265: 3851–3858.
Bruch, M.D., McKnight, C.J., and Gierasch, L.M. (1989) *Biochemistry* 28: 8554–8561.
Chakrabartty, A., Schellman, J.A., and Baldwin, R.L. (1991) *Nature* 351: 586–588.
Chou, M.M., and Kendall, D.A. (1990) *J Biol Chem* 265: 2873–2880.
Chou, P.Y., and Fasman, G.D. (1978) *Annu Rev Biochem* 47: 251–276.
Cornell, D.G., Dluhy, R.A., Briggs, M.S., McKnight, C.J., and Gierasch, L.M. (1989) *Biochemistry* 28: 2789–2797.
Dalbey, R.E. (1991) *Mol Microbiol* 5: 2855–2860.
Deber, C.M., Brandl, C.J., Deber, R.B., Hsu, L.C., and Young, X.K. (1986) *Arch Biochem Biophys* 251: 68–76.
Derman, A.I., Puziss, J.W., Bassford, Jr, P.J., and Beckwith, J. (1993) *EMBO* 12: 879–888.

Doud, S.K., Chou, M.M., and Kendall, D.A. (1993) *Biochemistry* 32: 1251–1256.
Driessen, A.J.M. (1992) *Trends Biochem Sci* 17: 219–223.
Duffaud, G., and Inouye, M. (1988) *J Biol Chem* 263: 10224–10228.
Emr, S.D., and Silhavy, T.J. (1983) *Proc Natl Acad Sci USA* 80: 4599–4603.
Engelman, D.M., and Steitz, T.A. (1981) *Cell* 23: 411–422.
Fikes, J.D., and Bassford, Jr, P.J. (1989) *J Bacteriol* 171: 402–409.
Fikes, J.D., Barkocy-Gallagher, G.A., Klapper, D.G., and Bassford, Jr, P.J. (1990) *J Biol Chem* 265: 3417–3423.
Francetic, O., Hanson, M.P., and Kumamoto, C.A. (1993) *J Bacteriol* 175: 4036–4044.
Geller, B.L. (1991) *Mol Microbiol* 5: 2093–2098.
Gennity, J., Goldstein, J., and Inouye, M. (1990) *J Bioenerg Biomemb* 22: 233–269.
Gething, M.-J., and Sambrook, J. (1992) *Nature* 355: 33–45.
Goldstein, J., Lehnhardt, S., and Inouye, M. (1990) *J Bacteriol* 172: 1225–1231.
Goldstein, J., Lehnhardt, S., and Inouye, M. (1991) *J Biol Chem* 266: 14413–14417.
Hardy, S.J.S., and Randall, L.L. (1991) *Science* 251: 439–443.
von Heijne, G. (1984) *J Mol Biol* 173: 243–251.
von Heijne, G. (1985a) In *Current Topics in Membranes and Transport*. Bronner, F. (ed.). New York: Academic Press, pp. 151–179.
von Heijne, G. (1985b) *J Mol Biol* 184: 99–105.
von Heijne, G. (1986) *J Mol Biol* 192: 287–290.
von Heijne, G. (1992) *J Mol Biol* 225: 487–494.
von Heijne, G., and Abrahmsén, L. (1989) *FEBS Lett* 244: 439–446.
von Heijne, G., and Blomberg, C. (1979) *Eur J Biochem* 97: 175–181.
Hikita, C., and Mizushima, S. (1992a) *J Biol Chem* 267: 4882–4888.
Hikita, C., and Mizushima, S. (1992b) *J Biol Chem* 267: 12375–12379.
Hoyt, D.W., and Gierasch, L.M. (1991) *Biochemistry* 30: 10155–10163.
Inouye, M., and Halegoua, S. (1980) *CRC Crit Rev Biochem* 7: 339–371.
Inouye, S., Soberon, X., Franceschini, T., Nakamura, K., Itakura, K., and Inouye, M. (1982) *Proc Natl Acad Sci USA* 79: 3438–3441.
Ito, K. (1992) *Mol Microbiol* 6: 2423–2428.
Jain, R.G., Rusch, S.L., and Kendall, D.A. (1994) *J Biol Chem* 269: 16305–16310.
Joly, J.C., and Wickner, W. (1993) *EMBO* 12: 255–263.
Jones, J.D., McKnight, C.J., and Gierasch, L.M. (1990) *J Bioenerg Biomemb* 22: 213–232.
Kadonaga, J.T., Plückthun, A., and Knowles, J.R. (1985) *J Biol Chem* 260: 16192–16199.
Keller, R.C.A., Killian, J.A., and de Kruijff, B. (1992) *Biochemistry* 31: 1672–1677.
Kendall, D.A., and Kaiser, E.T. (1988) *J Biol Chem* 263: 7261–7265.
Kendall, D.A., Bock, S.C., and Kaiser, E.T. (1986) *Nature* 321: 706–708.
Kendall, D.A., Doud, S.K., and Kaiser, E.T. (1990) *Biopolymers* 29: 139–147.
Kuhn, A., and Wickner, W. (1985) *J Biol Chem* 260: 15914–15918.
Kusters, R., Breukink, E., Gallusser, A., Kuhn, A., and de Kruijff, B. (1994) *J Biol Chem* 269: 1560–1563.
Laforet, G.A., and Kendall, D.A. (1991) *J Biol Chem* 266: 1326–1334.
Landry, S.J., and Gierasch, L.M. (1991) *Biochemistry* 30: 7359–7362.
Lehnhardt, S., Pollitt, S., and Inouye, M. (1987) *J Biol Chem* 262: 1716–1719.
Li, S.-C., and Deber, C.M. (1993) *J Biol Chem* 268: 22975–22978.
Luirink, J., High, S., Wood, H., Giner, A., Tollervey, D., and Dobberstein, B. (1992) *Nature* 359: 741–743.
Lyu, P.C., Sherman, J.C., Chen, A., and Kallenbach, N.R. (1991) *Proc Natl Acad Sci USA* 88: 5317–5320.
MacIntyre, S., Eschbach, M.-L., and Mutschler, B. (1990) *Mol Gen Genet* 221: 466–474.
Matsuyama, S., Fujita, Y., and Mizushima, S. (1993) *EMBO* 12: 265–270.
McKnight, C.J., Briggs, M.S., and Gierasch, L.M. (1989) *J Biol Chem* 264: 17293–17297.
McKnight, C.J., Rafalski, M., and Gierasch, L.M. (1991) *Biochemistry* 30: 6241–6246.
Mizushima, S., Tokuda, H., and Matsuyama, S. (1991) In *Methods in Cell Biology*. Tartakoff, A.M. (ed.). New York: Academic Press, pp. 107–146.
Nilsson, I., and von Heijne, G. (1992) *FEBS Lett* 299: 243–246.
Nishiyama, K., Mizushima, S., and Tokuda, H. (1993) *EMBO* 12: 3409–3415.
Nothwehr, S.N., and Gordon, J.I. (1990) *J Biol Chem* 265: 17202–17208.
Oliver, D.B. (1993) *Mol Microbiol* 7: 159–165.
Oliver, D.B., Cabelli, R.J., Dolan, K.M., and Jarosik, G.P. (1990) *Proc Natl Acad Sci USA* 87: 8227–8231.
Olsen, M.K., Rosey, E.L., and Tomich, C.-S.C. (1993) *J Bacteriol* 175: 7092–7096.
O'Neil, K.T., and DeGrado, W.F. (1990) *Trends Biochem Sci* 15: 59–64.
Osborne, R.S., and Silhavy, T.J. (1993) *EMBO* 12: 3391–3398.
Perlman, D., and Halvorson, H.O. (1983) *J Mol Biol* 167: 391–409.
Phillips, G.J., and Silhavy, T.J. (1992) *Nature* 359: 744–746.
Phoenix, D.A., de Wolf, F.A., Staffhorst, R.W.H.M., Hikita, C., Mizushima, S., and de Kruijff, B. (1993a) *FEBS Lett* 324: 113–116.
Phoenix, D.A., Kusters, R., Hikita, C., Mizushima, S., and de Kruijff, B. (1993b) *J Biol Chem* 268: 17069–17073.
Pugsley, A.P. (1993) *Microbiol Rev* 57: 50–108.
Puziss, J.W., Fikes, J.D., and Bassford, Jr, P.J. (1989) *J Bacteriol* 171: 2303–2311.
Puziss, J.W., Strobel, S.M., and Bassford, Jr, P.J. (1992) *J Bacteriol* 174: 92–101.
Rizo, J., Blanco, F.J., Kobe, B., Bruch, M.D., and Gierasch, L.M. (1993) *Biochemistry* 32: 4881–4894.
Rusch, S.L., and Kendall, D.A. (1992) *J Mol Biol* 224: 77–85.

Rusch, S.L., and Kendall, D.A. (1994) *J Biol Chem* 269: 124?248.
Rusch, S.L., Chen, H., Izard, J.W., and Kendall, D.A. (1994) *J Cell Biochem* 55: 209–217.
Sasaki, S., Matsuyama, S., and Mizushima, S. (1990) *J Biol Chem* 265: 4358–4363.
Schatz, P.J., and Beckwith, J. (1990) *Annu Rev Genet* 24: 215–248.
Schneider, G., Rõhlk, S., and Wrede, P. (1993) *Biochem Biophys Res Commun* 194: 951–959.
Shen, L.M., Lee, J.-I., Cheng, S., Jutte, H., Kuhn, A., and Dalbey, R.E. (1991) *Biochemistry* 30: 11775–11781.
Stader, J., Benson, S.A., and Silhavy, T.J. (1986) *J Biol Chem* 261: 15075–15080.
Stader, J., Gansheroff, L.J., and Silhavy, T.J. (1989) *Genes Devel* 3: 1045–1052.
Sung, C.Y., Gennity, J.M., Pollitt, N.S., and Inouye, M. (1992) *J Biol Chem* 267: 997–1000.
Tahara, Y., Murata, M., Ohnishi, S., Fujiyoshi, Y., Kikuchi, M., and Yamamoto, Y. (1992) *Biochemistry* 31: 8747–8754.
Taura, T., Baba, T., Akiyama, Y., and Ito, K. (1993) *J Bacteriol* 175: 7771–7775.
Thom, J.R., and Randall, L.L. (1988) *J Bacteriol* 170: 5654–5661.
Vlasuk, G.P., Inouye, S., Ito, H., Itakura, K., and Inouye, M. (1983) *J Biol Chem* 258: 7141–7148.
Vlasuk, G.P., Inouye, S., and Inouye, M. (1984) *J Biol Chem* 259: 6195–6200.
de Vrije, T., de Swart, R.L., Dowhan, W., Tommassen, J., and de Kruijff, B. (1988) *Nature* 334: 173–175.
de Vrije, G.J., Batenburg, A.M., Killian, J.A., and de Kruijff, B. (1990) *Mol Microbiol* 4: 143–150.
Wallace, B.A., Cascio, M., and Mielke, D.L. (1986) *Proc Natl Acad Sci USA* 83: 9423–9427.
Watson, M.E.E. (1984) *Nucl Acids Res* 12: 5145–5164.
Weiss, J.B., and Bassford, Jr, P.J. (1990) *J Bacteriol* 172: 3023–3029.
Wickner, W. (1979) *Annu Rev Biochem* 48: 23–45.
Wickner, W., Driessen, A.J.M., and Hartl, F.-U. (1991) *Annu Rev Biochem* 60: 101–124.